

# DU TEST DE VALIDITE D'UN AJUSTEMENT BASE SUR LES FONCTIONS DE L'ORDRE DES OBSERVATIONS

par R. SNEYERS

*Institut Royal Météorologique de Belgique*

## 1. Introduction.

Le contrôle de la validité de l'ajustement d'une loi de probabilité à une série d'observations se fait habituellement (cf. [1], p. 42 et 47) en utilisant le test de Kolmogorov-Smirnov ou le test en  $\chi^2$ , fondés tous deux sur les différences entre les fréquences observées et les fréquences ajustées correspondantes, cumulées dans le premier cas et non cumulées dans le second.

Il ne fait aucun doute qu'en raison de leur caractère global, l'un et l'autre de ces deux tests souffrent d'un manque de sensibilité. Le premier considère, en effet, la valeur extrême des différences entre les fréquences cumulées observées et les fréquences cumulées ajustées sans tenir compte de la valeur de la fréquence ajustée qui conduit à la différence maximale, tandis que le second exige le groupement des fréquences ajustées non cumulées de manière à atteindre un nombre théorique minimal de cinq. Il en résulte, en particulier, que ces méthodes ne permettent vraisemblablement pas d'apprécier correctement la compatibilité de la loi ajustée avec les valeurs extrêmes observées.

Le but de la présente note est de rappeler les éléments d'une méthode, préconisée par Gumbel (cf. [2], p. 52) sous la forme de courbes de contrôle, qui évite ces écueils, d'établir cette méthode successivement pour le cas d'une distribution à variable continue et celui d'une distribution à variable discrète et d'illustrer cette méthode par quelques exemples.

## 2. Généralités.

Si  $F$  désigne la fonction de distribution de la variable aléatoire quelconque  $x$  et si l'on effectue  $n$  observations indépendantes de cette variable, on sait (cf. par ex. [3], p. 94) que la fonction de distribution  $\phi_m$  de la  $m^{\text{e}}$  valeur, comptée à partir de la plus petite, est donnée par la formule :

$$\phi_m = \sum_{v=m}^n \binom{n}{v} F^v (1 - F)^{n-v}. \quad (1)$$

En particulier, les fonctions de distribution  $\phi_n, \phi_{n-1}, \phi_{n-2}, \dots$  de la plus grande valeur, de la deuxième plus grande valeur, de la troisième plus

grande valeur, ... s'obtiennent en arrêtant le développement du binôme de Newton  $[F + (1 - F)]^n$  respectivement à son premier terme, à son second terme, à son troisième terme, etc.

Il en résulte que si l'on désigne par  $F_m$  la valeur de  $F$  correspondant à la  $m^{\text{e}}$  valeur, on peut construire pour  $F_m$  un intervalle de confiance, avec le coefficient  $(1 - 2\alpha)$  où  $\alpha < 0,5$ , dont les limites s'obtiennent en résolvant par rapport à  $F$  les équations :

$$\phi_m = \alpha \quad \text{et} \quad \phi_m = 1 - \alpha. \quad (2)$$

Pour la résolution des équations (2), trois méthodes peuvent être retenues.

La première revient à noter avec Hald (cf. [3], p. 674) que, moyennant la transformation :

$$F = \frac{v_1 v^2}{v_2 + v_1 v^2}, \quad \text{avec } v_1 = 2m \text{ et } v_2 = 2(n - m + 1), \quad (3)$$

qui donne aussi :

$$v^2 = \frac{v_2}{v_1} \cdot \frac{F}{1 - F}, \quad (4)$$

la fonction (1) s'identifie avec la fonction de distribution du rapport de variance  $v^2$  (ou  $F$  de Snedecor), pour les degrés de liberté  $v_1$  et  $v_2$ .

A ce propos, on se souviendra, en effet, que la densité de probabilité  $f(v^2)$  de cette dernière distribution, qui peut s'écrire :

$$f(v^2) = \frac{\Gamma\left(\frac{v_1 + v_2}{2}\right)}{\Gamma\left(\frac{v_1}{2}\right) \Gamma\left(\frac{v_2}{2}\right)} \cdot (v_1)^{v_1/2} (v_2)^{v_2/2} (v^2)^{(v_1/2)-1} (v_2 + v_1 v^2)^{-(v_1+v_2)/2}, \quad (5)$$

pour  $0 \leq v^2 \leq \infty$

devient, moyennant la transformation (4) :

$$\frac{\Gamma\left(\frac{v_1 + v_2}{2}\right)}{\Gamma\left(\frac{v_1}{2}\right) \Gamma\left(\frac{v_2}{2}\right)} F^{(v_1/2)-1} (1 - F)^{(v_2/2)-1}, \quad \text{avec } 0 \leq F \leq 1, \quad (6)$$

puisque  $\frac{d v^2}{d F} = \frac{v_2}{v_1} (1 - F)^{-2}$ .

Il s'ensuit que si l'on pose  $v_1 = 2m$  et  $v_2 = 2(n - m + 1)$ , la fonction (6) devient :

$$\frac{\Gamma(n+1)}{\Gamma(m)\Gamma(n-m+1)} \cdot F^{m-1} (1-F)^{n-m}, \quad (7)$$

c'est-à-dire la dérivée de  $\phi_m$  par rapport à  $F$ .

On en déduit que la résolution des équations (2) peut se faire en déterminant les valeurs critiques de  $v^2$  pour le niveau  $2\alpha$  et pour les degrés de liberté  $v_1 = 2m$  et  $v_2 = (n - m + 1)$ , et en calculant les valeurs correspondantes de  $F_m$  au moyen de la relation (3).

La seconde méthode trouve son origine dans le fait que si l'on compte le rang  $k$  des valeurs observées à partir de la plus grande, on a :  $m = n - k + 1$  et la fonction (7) s'écrit :

$$\frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} \cdot F^{n-k} (1-F)^{k-1}, \quad (8)$$

ce qui, moyennant la transformation  $w = n(1 - F)$ , donne (cf. [4], p. 370-371), lorsque  $n$  tend vers l'infini :

$$\lim_{n \rightarrow \infty} \phi_{n-k+1} = \int_w^\infty \frac{w^{k-1}}{\Gamma(k)} e^{-w} dw. \quad (9)$$

On en conclut, avec Borgman [5], que si  $\phi \neq 0$ ,  $w$  possède une valeur finie lorsque  $n$  tend vers l'infini. Cette valeur finie peut être déduite de l'équation (9) avec l'aide de tables de la distribution gamma {par ex. [6] où  $p = k - 1$ ,  $u = w/\sqrt{k}$  et  $\lim_{n \rightarrow \infty} \phi_{n-k+1} = 1 - I(u, p)$ }, tandis que pour les valeurs finies de  $n$ , le calcul de  $w$  se fera en intégrant la fonction (8) au moyen de tables de la distribution bêta.

Des valeurs de  $w$  ont été calculées de cette manière par Borgman [5] pour  $k = 1, 2, 3, 4, 5$ , pour  $\phi = 0,005, 0,010, 0,025, 0,050, 0,100$  (0,100) 0,900, 0,950, 0,975, 0,990 et 0,995 et pour  $n = 1$  (1) 10, 12 (2) 20, 25, 30 (10) 60, 80, 100, 140 (40) 300, 350, 400, 500, 700, 1000, 2000 et  $\infty$ . Leur emploi permet de calculer les valeurs limites de  $F$  au moyen de la relation :

$$F = 1 - \frac{w}{n}. \quad (10)$$

La troisième méthode concerne les valeurs proches de la médiane pour lesquelles on retiendra avec Gumbel (cf. [2], p. 48) que lorsque  $n$  est suffi-

samment grand, la fréquence associée à la  $m^e$  valeur est asymptotiquement distribuée suivant une loi normale de moyenne  $p = m/(n + 1)$  et de variance  $p q/n$ , avec  $q = 1 - p$ .

### 3. Le test de validité. Courbes de contrôle.

#### a) Cas d'une distribution à variable continue.

Soit  $F(x)$  la fonction de distribution ajustée à la série des valeurs observées  $x_1 \leq x_2 \leq \dots \leq x_m \leq \dots \leq x_n$ . Le test de validité de l'ajustement consiste dans le choix d'un seuil de signification  $2\alpha$ , dans le calcul des limites de confiance  $\underline{F}_m$  et  $\overline{F}_m$  de  $F_m$  pour le coefficient  $(1 - 2\alpha)$  au moyen des équations (2) et dans l'acceptation ou le rejet de la valeur  $F(x_m)$  que l'ajustement effectué associe à la  $m^e$  valeur  $x_m$ , selon que cette valeur se trouve ou non à l'intérieur de l'intervalle  $(\underline{F}_m, \overline{F}_m)$ .

On peut également déterminer les valeurs  $\underline{x}_m$  et  $\overline{x}_m$  au moyen des équations :

$$F(\underline{x}_m) = \underline{F}_m \quad \text{et} \quad F(\overline{x}_m) = \overline{F}_m \quad (11)$$

et vérifier ensuite si  $x_m$  se trouve ou non à l'intérieur de l'intervalle  $(\underline{x}_m, \overline{x}_m)$ .

Les courbes de contrôle de la distribution  $F(x)$  s'obtiennent en procédant à une représentation graphique des intervalles  $(\underline{x}_m, \overline{x}_m)$  et en joignant entre eux les points qui délimitent ces intervalles.

La construction de ces courbes peut se simplifier si l'on note (cf. [2], p. 49) qu'au voisinage de la médiane la variable  $x_m$  est distribuée suivant une loi normale dont la moyenne  $\hat{x}_m$  est la solution de l'équation :

$$F(\hat{x}_m) = p = \frac{m}{n + 1}, \quad (12)$$

et dont l'écart-type  $\sigma(x_m)$  est donné par la relation :

$$\sigma(x_m) = \frac{\sqrt{p q}}{\sqrt{n} \cdot F'(\hat{x}_m)}, \quad (13)$$

où  $F'(x)$  est la densité de probabilité  $dF/dx$  de la variable  $x$ , avec  $q = 1 - p$ .

Dans ces conditions, si la variable  $x$  considérée est liée à une variable réduite  $z$  par une relation linéaire de la forme :

$$z = a(x - b) \quad (14)$$

on pourra, grâce à l'égalité :

$$\sigma(x_m) = [\sigma(z_m)]/a$$

calculer directement  $\sigma(x_m)$  à partir des  $\sigma(z_m)$  déterminés une fois pour toutes au moyen de la relation (13) où l'on a remplacé  $x_m$  par  $z_m$ .

En particulier, lorsque  $x$  est distribué suivant une loi normale ou une loi doublement exponentielle, on pourra calculer  $\sigma(x_m)$  en faisant appel aux tables de  $\sqrt{n} \cdot \sigma(z_m)$  publiées dans [2], p. 52 et 214.

Enfin, lorsqu'on procède à une représentation des observations et de la loi ajustée sur un papier à échelle fonctionnelle appropriée, les probabilités étant marquées en abscisse et les valeurs de la variable en ordonnée, l'équation (12) montre qu'il convient de placer les valeurs observées et les intervalles  $(x_m, \bar{x}_m)$  sur la parallèle à l'axe des ordonnées d'abscisse  $m/(n + 1)$ .

b) *Cas d'une distribution à variable discrète.*

Soient  $p(i)$ ,  $i = 0, 1, 2, \dots, N$ , les probabilités associées à la variable discrète  $i$  et soit  $F(x) = \sum_{i=0}^x p(i)$ ,  $x = 0, 1, 2, \dots, N$ , la fonction de distribution de cette variable avec  $F(N) = 1$ . Il est clair que dans ce cas la condition de compatibilité devient :

$$F(x_m - 1) < F_m < F(x_m),$$

sachant que  $F_m$  est l'une quelconque des valeurs comprises à l'intérieur de l'intervalle  $(F_m, \bar{F}_m)$ .

En d'autres termes, le rejet de l'ajustement n'aura lieu que si l'on a :

$$\bar{F}_m \leq F(x_m - 1) \quad \text{ou} \quad F_m \geq F(x_m), \quad (15)$$

tandis que dans les autres cas cet ajustement pourra être accepté.

Il reste à signaler que lors de l'application répétée du test de validité avec un seuil de signification  $\beta$ , dans l'éventualité où l'ajustement convient réellement, on doit s'attendre à voir apparaître en moyenne  $100 \beta \%$  de résultats significatifs. De plus, si sur  $N_1$  résultats indépendants on trouve

$k$  résultats significatifs, le résultat global ne peut être considéré comme significatif pour le niveau  $2\alpha$  que si l'on a :

$$\begin{aligned} \phi_{N_1-k+1}(1-\beta) &= (1-\beta)^{N_1} + N_1(1-\beta)^{N_1-1} \cdot \beta + \dots \\ &+ \binom{N_1}{k-1} (1-\beta)^{N_1-k+1} \cdot \beta^{k-1} > (1-2\alpha), \quad (16) \end{aligned}$$

puisque  $\phi_{N_1-k+1}(1-\beta)$  représente ici la probabilité d'obtenir au plus  $(k-1)$  résultats significatifs (de probabilité  $\beta$ ).

#### 4. Exemples.

Les exemples d'ajustements que nous avons considérés pour illustrer la théorie précédente sont relatifs à la distribution à Uccle des moyennes hivernales de la température de l'air (variable continue) et à celle des fréquences mensuelles des jours de précipitations mesurables (variable discrète).

a) *Les moyennes hivernales de la température de l'air à Uccle (1834-1963).*

La recherche de la loi de probabilité qui régit les moyennes hivernales de la température de l'air présente un intérêt particulier puisqu'elle comprend celle de la probabilité attachée au retour d'un hiver aussi rigoureux que celui qui vient d'être observé.

Les moyennes dont nous avons disposé sont celles qui ont été publiées dans [7]. Nous rappellerons à ce sujet que ces moyennes sont en réalité les sommes des moyennes mensuelles<sup>(1)</sup> de décembre, janvier et février (hiver climatologique), calculées à partir des extrêmes diurnes relevés à midi et exprimées en 0,1°C, ce qui équivaut à dire que nos moyennes hivernales sont données en tiers de 0,1°C. Ces données qui s'étendent, pour les hivers, de 1834 à 1956, ont été complétées par celles enregistrées de 1957 à 1963.

Comme l'analyse de la stabilité de la série entière avait fait ressortir l'existence d'une modification des propriétés statistiques de ces moyennes à partir de 1902 (cf. [7], p. 15), en même temps que le caractère homogène des séries partielles de 1834 à 1901 et de 1902 à 1956, une loi de probabilité a été ajustée à chacune des séries partielles.

(1) Quelques erreurs se sont glissées dans les chiffres publiés dans [7]. En particulier, il faut lire :

p. 20 : 1919, V : 144 au lieu de 185 ; 1920, I : 45 au lieu de 41 ; 1924, XII : 18 au lieu de 27 ;

p. 21 : 1953, I : 08 au lieu de 75.

De plus, les sommes trimestrielles et annuelles correspondantes doivent être corrigées d'autant.

Ces lois sont des lois normales et le calcul des paramètres a été effectué par la méthode de Gumbel (cf. [2], p. 38). Ces valeurs ont été indiquées au tableau I où, pour chaque série,  $n$  désigne l'effectif de la série,  $b = \bar{x}$  est la moyenne des observations,  $s$  est la racine carrée de l'estimation absolument correcte de la variance [ $s^2 = \Sigma (x - \bar{x})^2 / (n - 1)$ ] et  $\sigma_n$  est la moyenne quadratique des limites des quantiles d'ordre  $(n + 1)$  de la distribution normale réduite (cf. [2], p. 39).

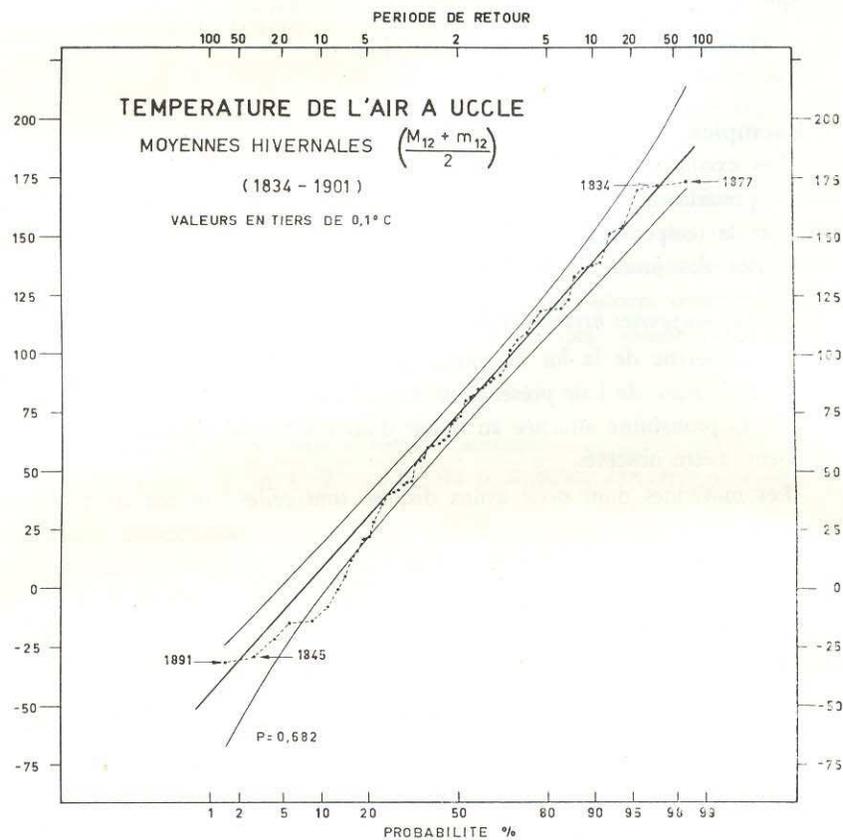


Fig. 1. — Loi normale ajustée, courbes de contrôle (en traits pleins) et série d'observations (en traits ponctués fins).

Il s'ensuit que l'ajustement s'obtient en liant la température hivernale  $x$  à la variable réduite  $z$  au moyen de la relation (14) ou encore de la relation :

$$x = b + \frac{1}{a} z, \quad (17)$$

ce qui permet de représenter la loi ajustée par une droite sur un papier à échelle fonctionnelle normale.

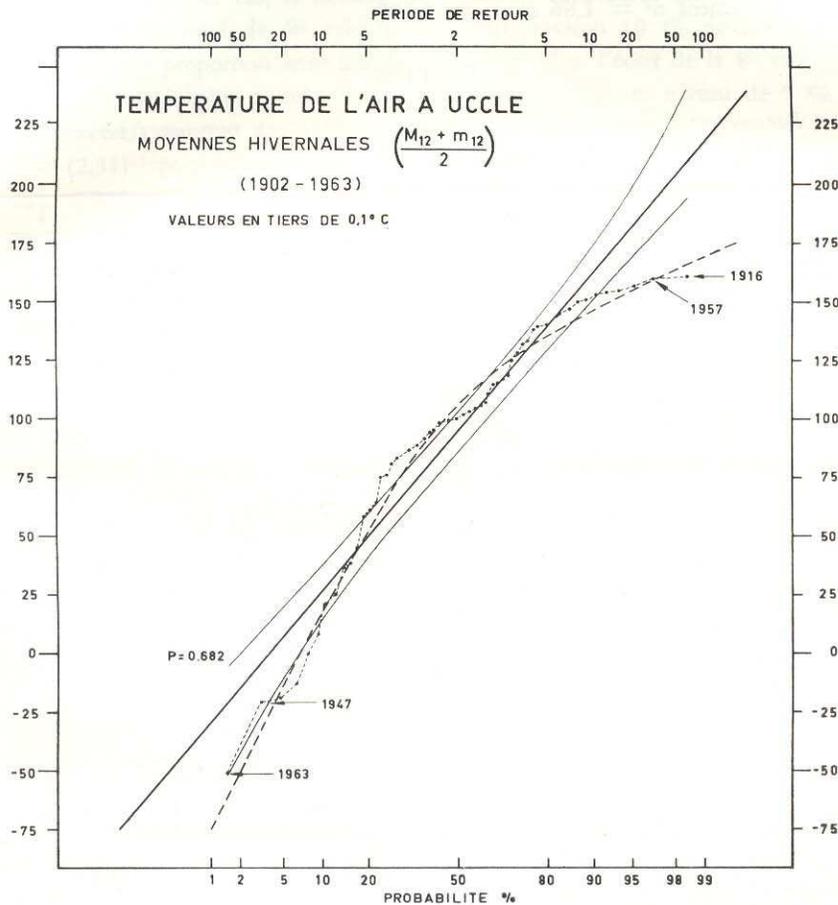


Fig. 2. — Loi normale ajustée, courbes de contrôle (en traits pleins), série d'observations (en traits ponctués fins) et loi normale corrigée (en traits ponctués épais).

Cette représentation graphique a été faite aux figures 1 et 2 en même temps que celle des observations et celle des courbes de contrôle. Celles-ci ont été établies pour le coefficient de confiance  $P = 0,682$  qui correspond à l'intervalle  $(-1, +1)$  de la variable réduite normale. Pour les valeurs proches de la médiane, les courbes ont été obtenues en joignant les extrémités des intervalles de confiance entourant les points d'abscisses  $F = 0,15, 0,20, 0,30, \dots, 0,80, 0,85$  de la droite ajustée, ces intervalles ayant été construits en utilisant les écarts-types donnés au tableau II. Pour les valeurs extrêmes

on a tiré de la table 2 de [5] les valeurs de  $w$  correspondant à  $\phi = 0,159$  et  $\phi = 0,841$ , avec  $n = 68$  et  $n = 62$ , valeurs qui sont dans les deux cas respectivement  $w = 1,86$  et  $w = 0,175$ .

TABLEAU I

*Paramètres des distributions normales ajustées aux moyennes hivernales de la température de l'air à Uccle*

	$n$	$b = \bar{x}$	$s$	$\sigma_n$	$\frac{1}{a} = \frac{s}{\sigma_n}$	$\frac{1}{a\sqrt{n}}$
1834 - 1901	68	73,46	47,88	0,94707	50,56	6,13
1902 - 1963	62	94,02	49,99	0,94310	53,00	6,73

On en déduit, grâce à la formule (10), pour  $n = 68$ ,  $\bar{F}_n = 0,9726$  et  $\bar{F}_n = 0,99743$  et pour  $n = 62$ ,  $\bar{F}_n = 0,9700$  et  $\bar{F}_n = 0,99178$ , tandis que les valeurs correspondantes pour  $\bar{F}_1$  s'obtiennent en faisant  $F_1 = 1 - F_n$ .

Il en résulte que les intervalles de confiance qui doivent entourer les points d'abscisses  $1/(n+1)$  et  $n/(n+1)$  sur les droites ajustées et qui correspondent à la plus petite et à la plus grande valeur de  $x$ , sont respectivement dans le premier cas :  $(-68, -24)$  et  $(170, 215)$ , et dans le second cas :  $(-53, -6)$  et  $(194, 241)$ , ce qui permet de compléter le tracé des courbes de contrôle.

TABLEAU II

*Ecarts-types des valeurs proches de la médiane (distribution normale)*

$\frac{1}{a\sqrt{n}} =$	1	6,13	6,73
$P = 0,50$	1,253	7,68	8,43
0,60	1,268	7,77	8,53
0,70	1,318	8,08	8,87
0,80	1,429	8,76	8,62
0,85	1,532	9,39	10,30

Les valeurs pour  $\frac{1}{a\sqrt{n}} = 1$  sont tirées de la table 2.1.6 dans [2], p. 52.

On constate ainsi que, dans le premier cas, les petites valeurs de la 5<sup>e</sup> à la 11<sup>e</sup> sont à l'extérieur des courbes de contrôle et que, dans le second cas,

il en est de même pour les petites valeurs de la 3<sup>e</sup> à la 6<sup>e</sup>, les valeurs proches de la médiane de la 12<sup>e</sup> à la 30<sup>e</sup> et les grandes valeurs de la 58<sup>e</sup> à la 62<sup>e</sup>.

Dans le premier cas, le nombre de 7 valeurs extérieures aux courbes de contrôle sur un total de 68 valeurs représente environ 10 % de l'effectif total pour une proportion attendue de 32 %. De plus, l'écart de la 8<sup>e</sup> valeur (— 8), qui est le plus grand, n'est pas significatif pour un niveau de 5 % (test bilatéral) puisque de  $F(-8) = 0,0537$  on tire, avec la formule (4),  $v^2 = (2,31)^{-1}$  pour  $v_1 = 16$  et  $v_2 = 122$  degrés de liberté, pour lequel on trouve  $0,025 < \phi < 0,050$ .

L'ajustement peut donc être accepté.

TABLEAU III

Valeurs de  $\phi$  pour quelques valeurs observées de la série des températures de la période 1902-1963

Rang	Valeur observée	F	$\phi$
4	— 13	0,02265	$\phi \cong 0,05$
17	81	0,4030	$0,975 < \phi < 0,99$
58	154	0,8712	$0,05 < \phi < 0,10$
59	155	0,8751	$0,025 < \phi < 0,05$
60	157	0,8828	$0,010 < \phi < 0,025$
61	160	0,8934	$0,005 < \phi < 0,010$
62	161	0,8969	$\phi < 0,005$

Dans le second cas, par contre, la proportion de valeurs extérieures aux courbes de contrôle atteint cette fois 41 %, ce qui est sensiblement supérieur aux 32 % attendus. De plus, le calcul de  $\phi$  pour les valeurs qui s'écartent le plus de la droite d'ajustement, et qui ont été indiqués au tableau III, montre que la 17<sup>e</sup> et les trois plus grandes valeurs sont significatives pour un niveau de 5 % (test bilatéral) et qu'en outre la plus grande valeur conserve même ce caractère pour un niveau de 1 %.

Il s'ensuit que cette fois la loi ajustée doit être rejetée.

Un autre ajustement a été effectué en corrigeant l'équation (14) de la loi précédente au moyen d'un terme quadratique par une méthode que nous avons déjà utilisée ailleurs (cf. [9]).

On obtient ainsi la relation :

$$z = \frac{x - 94,02}{53,0} + 1,2563 x - 232,88 \\ + 2,0201 \sqrt{(x - 125,38)^2 + 1.054,2},$$

ce qui donne, à la figure 2, la loi représentée par la branche d'hyperbole qui s'accorde mieux avec la série des observations.

La comparaison des deux résultats trouvés permet de mettre en évidence le caractère de la modification climatique intervenue au cours des plus récentes années. Cette comparaison a été faite en calculant les périodes de retour  $T = 1/F$  ou  $T = 1/(1 - F)$ , selon qu'on a  $F < 0,5$  ou  $F > 0,5$ , pour les valeurs de  $x$  indiquées au tableau IV. Elle montre que cette évolution s'est manifestée à la fois par une augmentation de la probabilité des hivers les plus froids ( $x < 0$ ) et des hivers doux ( $100 < x < 150$ ) et par une diminution de la probabilité des hivers froids ( $0 < x < 50$ ) et des hivers très doux ( $x > 150$ ).

TABLEAU IV

*Périodes de retour des températures moyennes de l'hiver*  
(en tiers de  $0,1^{\circ}\text{C}$ )

$x$	-50	-25	0	25	50	75	100	125	150	175
1834 - 1901	133	40,0	13,4	5,8	3,1	2,0	3,3	6,4	15,1	43,8
1902 - 1963	51,0	26,1	14,3	8,4	5,2	3,3	2,3	3,1	12,4	189

b) *Les fréquences mensuelles des jours de précipitations mesurables à Bruxelles-Uccle.*

Dans ce qui suit, on a considéré les nombres de jours de précipitations mesurables (au moins 0,1 mm) observés à Bruxelles de 1833 à 1889 et à Uccle de 1890 à 1960, soit au total durant une période de 128 années.

Comme dans une étude précédente (cf. [10], p. 76), l'ajustement d'une distribution normale aux fréquences mensuelles de la période 1891 à 1959 a déjà fourni un résultat satisfaisant, le même type d'ajustement a été effectué pour chacun des douze mois de l'année, et ce en utilisant les observations de la période 1833-1960.

On notera, à ce propos, que l'application à ces séries d'observations d'un test de stabilité a fait connaître dans plusieurs cas l'existence de variations significatives complexes, notamment au cours de la période 1890 à 1960, ainsi que des différences significatives entre les moyennes pour les périodes 1833-1889 et 1890-1960. Toutefois, comme les valeurs extrêmes se sont produites aussi bien au cours de la première que de la seconde période, il nous a paru utile de faire les ajustements en négligeant les constatations précédentes.

Ces ajustements ont été effectués en calculant la moyenne  $\bar{x}$  et l'écart-type  $s$  de la série des observations au moyen des relations :

$$\bar{x} = \frac{\sum x}{n} \quad \text{et} \quad s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}, \quad \text{avec} \quad n = 128,$$

et la probabilité  $F(x)$  attachée à la fréquence  $x$  a été calculée à partir des relations :

$$F(x) = \varphi(z) \quad \text{avec} \quad z = \frac{x + 0,5 - \bar{x}}{s}$$

$$\text{et} \quad \varphi(z) = \int_{-\infty}^z \frac{e^{-(z^2/2)}}{\sqrt{2\pi}} dz.$$

Les valeurs des paramètres  $\bar{x}$  et  $s$  des distributions ainsi définies ont été indiquées au tableau V.

TABLEAU V

*Fréquences mensuelles des jours de précipitations mesurables à Bruxelles-Uccle*  
Valeurs de la moyenne de  $\bar{x}$  et de l'écart-type  $s$   
des distributions normales ajustées

	J	F	M	A	M	J
$\bar{x}$	18,398	16,234	16,805	15,453	15,164	14,898
$s$	4,9666	4,8643	4,9977	5,3593	5,0459	4,6412
	J	A	S	O	N	D
$\bar{x}$	16,234	16,016	15,359	17,805	18,336	18,953
$s$	5,1870	5,0275	5,1140	4,6348	4,8282	5,0678

En vertu de ce qui précède, l'ajustement devra être rejeté dès que l'une des relations (15) est vérifiée, c'est-à-dire dès que l'on a :

$$\bar{T}_m \leq T(x_m) \quad \text{ou} \quad \underline{T}_m \geq T(x_m + 1), \quad (18)$$

sachant que :

$$T_m = \frac{1}{1 - F_m} \quad \text{et} \quad T(x_m) = \frac{1}{1 - F(x_m - 1)}$$

pour  $F_m$  et  $F(x_m - 1) > 0,5$ .

Par ailleurs, étant donné que les tables de Borgman fournissent les valeurs de  $w = n(1 - F)$ , on peut déduire  $T_m$  directement de ces tables au moyen de la formule  $T_m = n/w$ .

En particulier, pour  $n = 128$ , on trouve, pour les cinq plus grandes valeurs, les valeurs limites de  $T_m$  données au tableau VI, tandis que les périodes de retour  $T(x_m)$  et  $T(x_m + 1)$ , que les ajustements effectués associent aux cinq plus grandes valeurs des séries étudiées, se trouvent au tableau VII.

TABLEAU VI  
Valeurs limites de  $T_{n-k+1}$  pour  $n = 128$

$\phi =$	0,025	0,050	0,100	0,900	0,950	0,975
$k = 1$	35,3	43,4	56,1	1219	2510	5120
2	23,4	27,4	33,3	240	359	527
3	18,1	20,7	24,4	115	156	205
4	15,0	16,9	19,5	72,7	92,8	116
5	12,8	14,3	16,3	52,2	64,3	78,0

On constate ainsi qu'au tableau VII.A, aucune valeur ne dépasse les valeurs correspondantes les plus élevées du tableau VI et qu'au contraire, au tableau VII.B, sept valeurs sont inférieures à au moins une des valeurs correspondantes les plus basses du tableau VI. Plus précisément, on note pour  $k = 1$  et 2, une valeur de  $T(x_m)$  significativement trop faible pour le niveau 0,025 et trois valeurs trop faibles pour le niveau 0,100 et, pour  $k = 3$ , une valeur trop faible pour le niveau 0,100 (tests unilatéraux).

Comme on peut admettre ici qu'il s'agit de douze essais indépendants, ces résultats ne pourront être globalement significatifs au niveau 0,05 que si l'on a :

$$\phi_{N_1-k+1} (1 - \beta) > 0,95, \quad (19)$$

pour  $N_1 = 12$ ,  $k = 1$  et  $\beta = 0,025$ , ou pour  $N_1 = 12$ ,  $k = 3$  et  $\beta = 0,100$ .

En réalité, on trouve dans le premier cas, grâce aux tables de Borgman, que  $k = 1$  et  $w = n [1 - (1 - \beta)] = 0,30$  correspondent à  $0,7 < \phi < 0,8$  et, dans le second cas, que  $k = 3$  et  $w = 1,2$  correspondent à  $0,8 < \phi < 0,9$ . Il s'ensuit que le rejet global des ajustements ne s'impose pas.

Si l'on ajoute que le même résultat a été obtenu pour les petites valeurs et qu'en outre les périodes de retour observées des valeurs proches de la médiane s'écartent peu des valeurs théoriques, on en conclut que les ajustements effectués peuvent être acceptés dans toute leur étendue.

TABLEAU VII

Fréquences mensuelles des jours de précipitations mesurables à Uccle  
Périodes de retour associées aux plus grandes  $m^e$  valeurs  
( $m = n - k + 1$ )

A. — Valeurs de  $T(x_m)$ 

$k$	J	F	M	A	M	J
1	29,9	57,4	104	134	445	89,5
2	29,9	35,2	61,8	81,4	49,4	89,5
3	19,4	35,2	38,2	50,9	31,1	31,3
4	19,4	22,4	24,4	21,9	31,1	31,3
5	19,4	22,4	16,2	21,9	13,7	19,7
$k$	J	A	S	O	N	D
1	111	89,5	68,1	33,0	22,0	53,5
2	67,0	54,0	42,2	20,6	14,5	21,8
3	41,8	33,8	27,1	20,6	14,5	21,8
4	18,0	33,8	18,0	13,5	14,5	21,8
5	18,0	21,9	18,0	13,5	14,5	14,7

B. — Valeurs de  $T(x_m + 1)$ 

$k$	J	F	M	A	M	J
1	47,7	97,3	181	228	843	161
2	47,7	57,4	104	134	81,1	161
3	29,9	57,4	61,8	81,4	49,4	51,9
4	29,9	35,2	38,2	32,9	49,4	51,9
5	29,9	35,2	24,4	32,9	20,3	31,3
$k$	J	A	S	O	N	D
1	190	154	114	54,9	<b>34,7</b>	88,1
2	111	89,5	68,1	33,0	<b>22,0</b>	33,6
3	67,0	54,0	42,2	33,0	22,0	33,6
4	27,0	54,0	27,1	20,6	22,0	33,6
5	27,0	33,8	27,1	20,6	22,0	21,9

Les valeurs en italique et en grasses sont significatives pour des niveaux respectifs de 0,100 et de 0,025 (tests unilatéraux).

## 5. Conclusions.

En résumé, on a vu que le test de validité d'un ajustement basé sur les fonctions de l'ordre des observations permet non seulement d'établir avec une grande sûreté la compatibilité ou l'incompatibilité de l'ajustement avec la série des observations, mais aussi de déterminer les régions du domaine de définition de la variable pour lesquelles l'ajustement ne convient pas.

Accessoirement, cette méthode a permis, dans le cas des températures hivernales à Uccle, de déterminer avec précision les caractères du changement climatique qui est intervenu depuis le début du siècle, tandis qu'elle s'est révélée utile dans le cas des fréquences mensuelles des jours de précipitations mesurables à Uccle pour lever le doute quant à la validité globale des ajustements effectués.

## Bibliographie.

- [1] SIEGEL, S. : *Non Parametric Statistics for the Behavioral Sciences*, Mc Graw Hill, New York, 1956.
- [2] GUMBEL, E.J. : *Statistics of Extremes*, Columbia University Press, New York, 1958.
- [3] HALD, A. : *Statistical Theory with Engineering Applications*, Wiley, New York, 1955.
- [4] CRAMER, H. : *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1958.
- [5] BORGMAN, L.E. : The Frequency Distribution of Near Extremes, *Journ. Geoph. Res.*, 66, 10, pp. 3295-3307, 1961.
- [6] PEARSON, K. : *Tables of the Incomplete Gamma Function*, Cambridge University Press, Cambridge, 1957.
- [7] SNEYERS, R. : Connexions thermiques entre saisons consécutives à Bruxelles-Uccle. *Inst. R. Mét. Belg., Pub. B*, 23, 1958.
- [8] HALD, A. : *Statistical Tables and Formulas*, Wiley, New York, 1955.
- [9] SNEYERS, R. : On a special Distribution of Maximum Values, *Month. Weather Rev.*, 88, 2, pp. 66-69, 1960.
- [10] SNEYERS, R. : Sur la probabilité des sécheresses à Uccle (Belgique) et son influence dans la répartition statistique de la cote udométrique, *Ass. Int. Hydrol. Scient. Pub. n° 51, Commission des Eaux de surface*, pp. 72-80, 1960.